点群データを対象とした物体検出手法の視覚的説明に関する研究

TR22012 三原一真　　指導教授：藤吉弘亘

1. はじめに

自動運転システムを実現するために，周囲の障害物を検知する物体検出は重要なタスクである．自動運転システムのセンサには，RGB カメラ，LiDAR センサ，Radar センサが用いられる．LiDAR センサはレーザ光を照射し，障害物などの周囲の物体から跳ね返ってきた反射光を受光し，ToF 原理により 3 次元座標を取得する．取得した 3 次元座標は点群データとして，セマンティックセグメンテーションや物体検出などのタスクに利用される．点群データを入力とする物体検出手法として，PointPillars[1] がある．PointPillars は，点群データを Pillar と呼ばれる xy 平面に垂直な格子状の空間に分割し，局所的な特徴抽出をすることで，疑似画像に変換している．疑似画像に変換することで，畳み込み処理の適用が容易となり，高速な推論が可能となる．一方で，点群ベースの手法の問題点として，入力した点群データに対する予測処理過程はブラックボックスとして扱われており，人間による判断根拠の判別が困難であるという点がある．そこで本研究では，点群データを入力とする物体検出モデルの視覚的説明の実現を目指す．

2. 関連研究

本節では，点群データを入力とする物体検出手法の PointPillars と，画像データを入力とする判断根拠の視覚的説明手法の ODAM について述べる．

2.1. PointPillars

PointPillars[1] は，Pillar feature Net，Backbone，Detection head から構成される．PointPillars のネットワーク構造を図 1 に示す．Pillar Feature Net は，局所的な特徴を抽出するモジュールである．まず，3 次元座標と反射強度で構成する点群データ (x, y, z, r) を擬似画像へ変換する処理を行う．入力された点群データは，Pillar と呼ばれる xy 平面に垂直に定義される領域に分割され，(D, P, N) の 3 次元のテンソルに変換される．ここで，D は点の次元数，P は Pillar の数，N は Pillar 内の点の数である．Pillar 内の点群データは，Pillar 中心からのオフセットと Pillar 内の点群データの平均値から点の次元数を $D = 9$ に変換し，PointNet へ入力することで (C, P, N) の 3 次元テンソルとなる．ここで，C はチャンネル数である．次に，N 方向の最大値を取ることにより (C, P) の 2 次元テンソルへ変換し，Pillar を元の位置に戻すことで，(C, h, w) の Birds-Eye View の擬似画像となる．ここで，h, w は画像の縦と横の画素数である．Backbone では，Pillar Feature Net で特徴抽出した疑似画像に対して，大局的な特徴抽出を行う．Backbone は，3 層のストライドが 2 の畳み込み層から構成されており，特徴マップのスケールを変更して特徴抽出を行う．その後，それぞれの特徴マップに対して Deconvolution によりアップサンプリングした後，連結して出力する．Detection head では，Backbone の出力から，3D Bounding Box のパラメータとクラススコアの予測結果を出力する．予測結果に対して，Non-Muximum Suppression(NMS) を行い，冗長な検出結果を削除する．

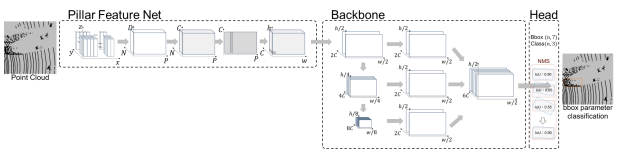


図 1: PointPillars のネットワーク構造

2.2. ODAM

ODAM[2] は，物体検出モデルの中間特徴の勾配情報から Attention Map を作成する．ODAM は，中間特徴に対する Attention Map を Bounding Box の 4 つのパラメータとクラススコアを逆伝播して求める．そして，中間特徴の要素ごとに Attention Map の最大値を取り，1 つの Attention Map にする．ODAM は，検出した Bounding Box に対して個々の Attention Map を求めることができる．

3. 提案手法

本研究では，点群データを入力とする物体検出手法の視覚的説明の実現を目的とする．提案手法はである Grad ODAM は，点群データを物体検出手法である PointPillars の Pillar 特徴による疑似画像に視覚的説明手法である ODAM を導入し，さらに Guided BackPropagation による Attention を統合することで，点群データに対する視覚的説明の獲得を目指す．提案手法を図 2 に示す．

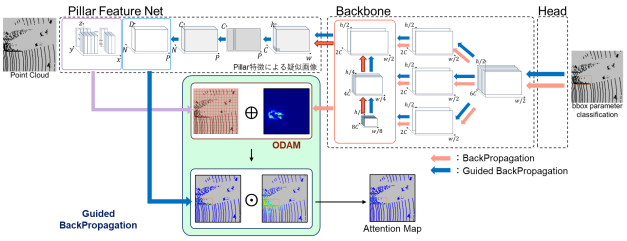


図 2: 提案手法のネットワーク構造

3.1. PointPillars への ODAM の導入

本研究では，PointPillars が，点群データを Birds-Eye View の疑似画像へ変換して特徴抽出する処理に着目し，疑似画像に対する Attention Map を求める．点群データを入力する際に Pillar Feature Net の出力である Pillar 特徴に対して，ODAM を用いて Attention Map を獲得する．次に，Pillar Feature Net のインデックス情報を利用して，Attention Map の値を対応する Pillar 内部の点群データに Attention として付与する．この場合，Attention を付与された点群データは，Pillar 内部で同値の Attetnion となる．

3.2. Guided ODAM

ODAM によって求めた Attention Map は，図 2 の Pillar Feature Net の最初の処理で分割した Pillar に対する Attention であるため，点ごとの視覚的説明ではない．そこで，逆伝播の際に，負の勾配に ReLU 関数を使用して 0 へ変換する Guided BackPropagation の Attention を統合する．これにより，点ごとに異なる Attention を付与することが可能となる．提案する Guided ODAM の式を式 (1) に示す．

$$A_{GuidedODAM} = (A_{ODAM} + g_{input})^2 \qquad (1)$$

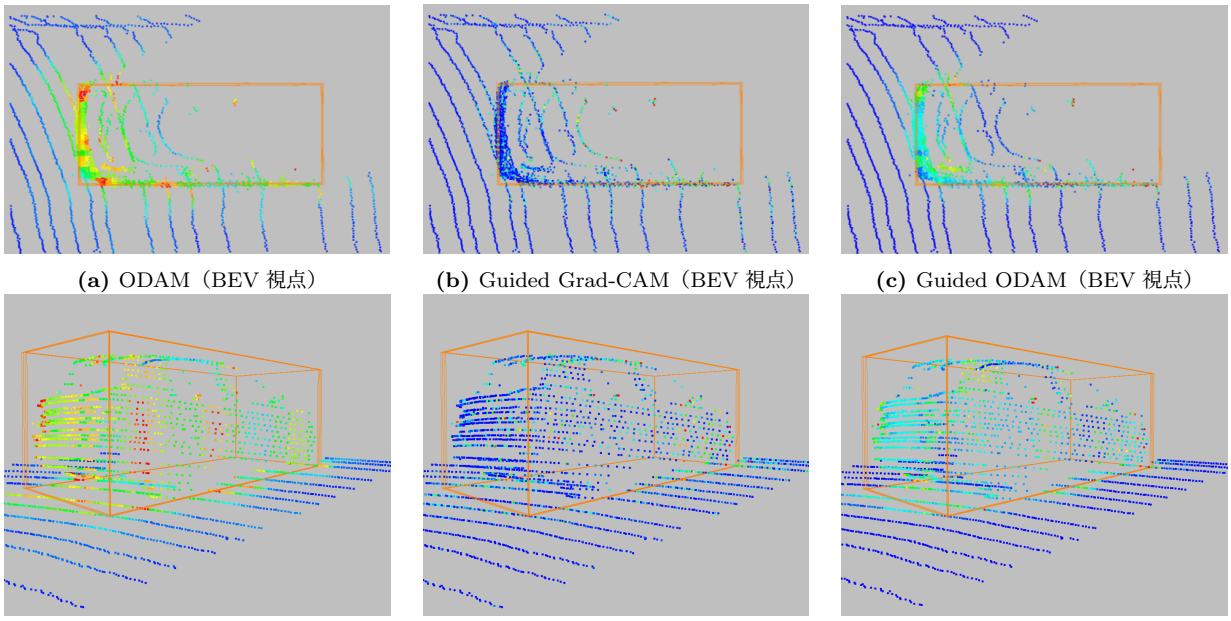
ここで A_{ODAM} は，ODAM の Attention を付与した点群データ，g_{input} は Guided BackPropagation のによるモデルの入力データの勾配情報である．Attention を付与した点群データに，Guided BackPropagation による勾配情報を加算し，2 乗することで Bounding Box 外の低い値の Attention を抑制する．これにより，Bounding Box 内部に高い値の Attention が発生し，点ごとの視覚的説明が可能な Attention Map を求めることができる．

4. 評価実験

本実験では，提案手法に導入する視覚的説明手法を，ODAM，Guided Grad-CAM，Guided ODAM に変更し，それぞれの視覚的説明手法を比較する．

4.1. KITTI dataset

本実験では，KITTI dataset[5] を使用する．KITTI dataset は，自動車に取り付けられた LiDAR センサや RGB カメラ等のセンサから作成された実環境データセットである．KITTI 3D object detection benchmark は，Car，Cyclist，Pedestrian の 3 クラスに対して Easy，Moderate，Hard



(a) ODAM（BEV 視点） **(b)** Guided Grad-CAM（BEV 視点） **(c)** Guided ODAM（BEV 視点）

(d) ODAM **(e)** Guided Grad-CAM **(f)** Guided ODAM

図 3: 視覚的説明手法ごとの定性的評価の比較

の 3 段階の難易度での識別精度の比較を行うベンチマークである．KITTI dataset の学習用データを分割し，3,712 枚，評価に 3,769 枚のデータを使用する．

4.2. 実験概要

本実験ではエポック数を 80，ミニバッチサイズを 8，初期学習率を 0.003，最適化手法を Adam に設定して学習する．

4.3. 評価指標

物体検出の視覚的説明において，Bounding Box 内部の物体に対して局所的に Attention を付与することで，物体検出モデルに対する説明性を可視化できる．本研究では，各予測結果に対する Attention Map 全体の Attention の総和と Bounding Box 内部の Attention の総和の割合の平均をクラスごとに計算することで，Attention Map の Bounding Box に対する局所性を求める．評価指標の計算式を式 (2) に示す．

$$mAPB = \frac{1}{M}\sum_{M}\frac{\sum^{n} Attention_{bbox}}{\sum^{N} Attention_{Map}} \tag{2}$$

M は Attention Map の枚数，n は Bounding Box 内の点群数，$Attention_{bbox}$ は Bounding Box 内の点群データの Attention，$Attention_{Map}$ は Attention Map のすべての点群データの Attention である．

4.4. 視覚的説明の評価結果

視覚的説明の定量的的評価を表 1 に示す．表 1 より，提案手法である Guided ODAM がすべてのクラスに対して最も高い精度であることが確認できる．これは，予測結果の Bounding Box 内部に高い値の Attention が発生し，Bounding Box 外に発生する Attention が抑制されている結果だと考えられる．

表 1: 視覚的説明における定量的評価の比較

手法	Car	Cyclist	Pedestrian
ODAM	71.39	42.65	53.27
Guided Grad-CAM	64.81	57.68	60.32
Guided ODAM	**80.56**	**68.00**	**73.69**

次に，定性的評価を図 3 に示す．図 3(a)，図 3(d) より，ODAM は，Bounding Box 内部が全体的に高い値の Attention が発生し，高さ方向に Attention を確認したときに点が同じ色となっていることが確認できる．これは，ODAM が Pillar 内部の特徴を点群データに付与した結果だと考えられる．そして，Bounding Box の外部にも高い値の Attention を確認できる．図 3(b)，図 3(e) より，Guided Grad-CAM は，Bounding Box 内にまばらに高い値の Attention を確認できる．また，Bounding Box の外部にも高い値の Attention を確認できる．図 3(c)，図 3(f) を見ると Guided ODAM は，Bounding Box 内部が全体的に高い値の Attention が発生し，高さ方向で Attention が変化していることを確認できる．これは，Guided BackPropagation による点ごとの視覚的説明によるものであると考えられる．また，ODAM と Guided Grad-CAM と比較すると，Bounding Box の外部に発生する Attention が低い値であることが確認できる．

4.5. おわりに

本研究では，点群データを入力とする物体検出手法に対する視覚的説明を行う手法を提案した．評価実験では，Guided ODAM は，定量的評価において最も評価指標の値が高く，定量的評価において点ごとの視覚的説明，Bounding Box 外の Attention の抑制による説明性の向上を確認した．今後は，ベースとなる物体検出手法や視覚的説明手法の変更を検討し，さらなる視覚的説明の向上を図る．

参考文献

[1] Alex H. Lang, Sourabh Vora, Holger Caesar, Lubing Zhou, Jiong Yang, Oscar Beijbom. PointPillars: Fast Encoders for Object Detection from Point Clouds. CVPR,2019.

[2] Chenyang, ZHAO and Chan, Antoni B. ODAM: Gradient-based Instance-Specific Visual Explanations for Object Detection. ICLR,2023.

[3] Selvaraju, Ramprasaath R and Cogswell, Michael and Das, Abhishek and Vedantam, Ramakrishna and Parikh, Devi and Batra, Dhruv. Grad-cam: Visual explanations from deep networks via gradient-based localization. ICCV,2017.

[4] Petsiuk, Vitali and Das, Abir and Saenko, Kate. Rise: Randomized input sampling for explanation of black-box models. arXiv,2018.

[5] Geiger, Andreas, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. CVPR,2022.

研究業績

[1] 三原一真　等，”Self-Attention を用いた PointPillars による 3 次元物体検出の高精度化”　画像センシングシンポジウム，2022.